

Bayes linear covariance matrix adjustment for multivariate dynamic linear models

Darren J Wilkinson * Michael Goldstein

University of Durham, UK

Abstract

A methodology is developed for the adjustment of the covariance matrices underlying a multivariate constant time series dynamic linear model. The covariance matrices are embedded in a distribution-free inner-product space of matrix objects which facilitates such adjustment. This approach helps to make the analysis simple, tractable and robust. To illustrate the methods, a simple model is developed for a time series representing sales of certain brands of a product from a cash-and-carry depot. The covariance structure underlying the model is revised, and the benefits of this revision on first order inferences are then examined.

Keywords: BAYES LINEAR METHODS; COVARIANCE MATRIX ESTIMATION; DYNAMIC LINEAR MODELS; EXCHANGEABILITY.

1 Introduction

Bayesian covariance matrix estimation is a notoriously difficult problem. For large matrices in particular, specification of beliefs about the distribution of the matrix is particularly hard. Distributional Bayesian approaches to the problem have tended to make use of restrictive conjugate prior assumptions in order to fix the prior distribution with as few hyper-parameters as possible. More recently, Leonard and Hsu (1992) have weakened the distributional assumptions required. However, specification within their formulation is very difficult, and, for large matrices, the computational problems are considerable. Specification is somewhat easier using the approach of Brown, Le, and Zidek (1994), but the other problems remain. In Wilkinson and Goldstein (1995), we outline a new approach to covariance matrix adjustment, exploiting second order exchangeability specifications, and a geometric space where random matrices live naturally.

*E-mail: d.j.wilkinson@durham.ac.uk — WWW: <http://fourier.dur.ac.uk:8000/djw.html> — For information on [B/D], go to <http://fourier.dur.ac.uk:8000/stats/bd/>

Covariance matrix adjustment for dynamic linear models is reviewed in West and Harrison (1989). For multivariate time series, the observational covariance matrix can be updated for a class of models known as *matrix normal models* using a simple conjugate prior approach. However, the distributional assumptions required are extremely restrictive, and it is difficult to learn about the covariance matrix for the updating of the state vector.

In this paper, we apply our approach to develop a methodology for the revision of the underlying covariance structures for a dynamic linear model, free from any distributional restrictions, using Bayes linear estimators for the covariance matrices based on simple quadratic observables. We do this by constructing an inner-product space of random matrices containing both the underlying covariance matrices and observables predictive for them. Bayes linear estimates for the underlying matrices follow by orthogonal projection.

We illustrate the method with data derived from the weekly sales of six leading brands of shampoo from a medium sized cash-and-carry depot. The sales are modelled taking into account the underlying demand and competition effects, and the covariance structure over the resulting dynamic linear model is adjusted using the weekly sales data.

2 The dynamic linear model

2.1 The general model

Let $\mathbf{X}_1, \mathbf{X}_2, \dots$ be an infinite sequence of random vectors, each of length r , such that $\mathbf{X}_t = (X_{1t}, X_{2t}, \dots, X_{rt})^T$. These vectors represent the observations at each time point. Suppose that we model the relationships between these vectors in the following way.

$$\mathbf{X}_t = F^T \boldsymbol{\Theta}_t + \boldsymbol{\nu}_t \quad (1)$$

$$\boldsymbol{\Theta}_t = G \boldsymbol{\Theta}_{t-1} + \boldsymbol{\omega}_t \quad (2)$$

Our prior second-order specification is as follows:

$$E(\boldsymbol{\nu}_t) = E(\boldsymbol{\omega}_t) = 0, \text{Var}(\boldsymbol{\Theta}_0) = \Sigma, \text{Var}(\boldsymbol{\nu}_t) = V, \text{Var}(\boldsymbol{\omega}_t) = W, \quad \forall t \quad (3)$$

$$\text{Cov}(\boldsymbol{\Theta}_s, \boldsymbol{\nu}_t) = \text{Cov}(\boldsymbol{\Theta}_s, \boldsymbol{\omega}_t) = \text{Cov}(\boldsymbol{\nu}_s, \boldsymbol{\omega}_t) = 0 \quad \forall s, t, \quad \text{Cov}(\boldsymbol{\omega}_s, \boldsymbol{\omega}_t) = \text{Cov}(\boldsymbol{\nu}_s, \boldsymbol{\nu}_t) = 0 \quad \forall s \neq t \quad (4)$$

where we adopt the usual conventions $\text{Cov}(\mathbf{A}, \mathbf{B}) = E(\mathbf{AB}^T) - E(\mathbf{A})E(\mathbf{B}^T)$ and $\text{Var}(\mathbf{A}) = \text{Cov}(\mathbf{A}, \mathbf{A})$. The *state vector*, $\boldsymbol{\Theta}_t$ is p dimensional, and the $p \times r$ and $p \times p$ dimensional matrices, F and G are assumed to be known. This is a second-order description of the (constant) multivariate time series dynamic linear model (DLM) described in West and Harrison (1989). We make no distributional assumptions for any of the components in the model. In this paper, we describe ways to learn about V and W from data. West

and Harrison (1989) (Chapter 15) give a conjugate prior solution to the problem of learning about V for a class of these models known as *matrix normal models*, if one is prepared to make the necessary distributional assumptions. However methods for learning about the matrix W tend primarily to be *ad hoc*.

2.2 Example

To illustrate our approach, we consider a simple locally constant model for the sales of 6 leading brands of shampoo from a medium sized cash-and-carry depot. As above $\mathbf{X}_1, \mathbf{X}_2, \dots$ is our sequence of random vectors, each of length 6, such that $\mathbf{X}_t = (X_{1t}, X_{2t}, \dots, X_{6t})^T$. The component X_{it} represents the (unknown) sales of brand i at time t . The vectors of sales are modelled as follows

$$\mathbf{X}_t = \boldsymbol{\Theta}_t + \boldsymbol{\nu}_t \quad \forall t \quad (5)$$

where

$$\boldsymbol{\Theta}_t = \boldsymbol{\Theta}_{t-1} + \boldsymbol{\omega}_t \quad \forall t \quad (6)$$

Prior beliefs are given by (3) and (4). Here we are assuming the process to be locally constant, but with different underlying demands for each of the components of the series. This is a simple model, with no seasonal component, chosen to illustrate our methodology, and would be unrealistic if there were noticeable trends within any of the components of the series. However, for high dimensional time series with no obvious trends, it is often the case that, provided the covariance structure is appropriate, many of the interesting features of the series can be captured using just such a model. To this end we introduce covariances between components of the state vector and also for the way demand changes over time, and for the way observations vary from the underlying demand. A more detailed treatment of multivariate sales forecasting within a fully specified Bayesian framework is given by Queen, Smith, and James (1994) and Queen (1994) who consider the problem of developing a dynamic model for multivariate sales, and the development of a prior distribution with sufficient flexibility to capture the effects of market interaction.

The second-order DLM requires the following quantifications. Firstly, the F and G matrices must be specified. Then, *a priori* specifications are needed for the expectation of the initial state vector, $\boldsymbol{\mu}_0 = E(\boldsymbol{\Theta}_0)$. Finally, we must specify the matrices $\Sigma = \text{Var}(\boldsymbol{\Theta}_0)$, $V = \text{Var}(\boldsymbol{\nu}_t)$, $W = \text{Var}(\boldsymbol{\omega}_t) \forall t$.

In our example the specification for the mean vector was

$$E(\boldsymbol{\Theta}_0) = (10, 9, 9, 8, 8, 7)^T \quad (7)$$

The following specifications were made for the covariance matrices, using exchangeability judgements concerning way the observations vary from their means. Using the notation A_{ij} for the $(i, j)^{th}$ element of the matrix A , we have

$$\Sigma_{ii} = 9 \quad \forall i, \quad \Sigma_{ij} = 3 \quad \forall i \neq j, \quad (8)$$

$$W_{ii} = 4 \forall i, W_{ij} = 1 \forall i \neq j \quad (9)$$

$$V_{ii} = 36 \forall i, V_{ij} = -4 \forall i \neq j \quad (10)$$

In truth, there is perhaps more symmetry in these specifications than is really appropriate, but specification is hard, and viewing variation in the sales of the various shampoos as second-order exchangeable greatly reduces the number of specifications which we have to make over the second order structure, and will allow further exchangeability modelling to simplify the fourth order specifications in later sections.

Notice however that many aspects of the underlying mechanisms have been captured by these specifications. In this model, Θ_t represents the vector of demands at time t . From the positive correlations in $\text{Var}(\Theta_0)$, if the mean of one product turned out to be higher than anticipated, we would revise upwards beliefs about the means of the other products. Also, the positive correlations within $\text{Var}(\omega_t)$ indicate that there is a common component to the demands, whilst the negative correlations within $\text{Var}(\nu_t)$ indicate that brands are competing, and tend to succeed at the expense of each other.

2.3 Bayes linear analysis

We take a Bayes linear approach to subjective statistical inference, making expectation (rather than probability) primitive. An overview of the methodology is given in Farrow and Goldstein (1993), in which the emphasis is on learning about means. With the second-order specification that we have made, we may use sales data to carry out an analogous Bayes linear analysis which will be informative for the mean of future observations. However, we will not learn about the covariance matrices $W = \text{Var}(\omega_t)$ or $V = \text{Var}(\nu_t)$. In this paper, we describe how such learning may take place.

3 Quadratic products

3.1 Exchangeable decomposition of unobservable products

For the matrix $A = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$, we define

$$\text{vec}A = (\mathbf{a}_1^T, \mathbf{a}_2^T, \dots, \mathbf{a}_n^T)^T \quad (11)$$

For the general DLM outlined in Section 2.1, we form the quadratic products of ω_t and ν_t , namely $\text{vec}(\omega_t \omega_t^T)$ and $\text{vec}(\nu_t \nu_t^T)$. We view $\text{vec}(\omega_t \omega_t^T)$ and $\text{vec}(\nu_t \nu_t^T)$ to be second-order exchangeable over t . By this, we mean that our second-order beliefs over the vectors of quadratic products of residuals will remain invariant under the action of an arbitrary permutation of the t index (this is what we mean when describing a DLM as *constant*). From the second-order exchangeability representation theorem (Goldstein 1986), we may represent an element of a second-order exchangeable collection of vectors as the sum of a

mean vector, common to all elements, and a residual vector, uncorrelated with the mean vector and all other residual vectors. We may apply this representation to $\mathbf{vec}(\boldsymbol{\omega}_t \boldsymbol{\omega}_t^T)$, and then re-write the representation in matrix form as

$$\boldsymbol{\omega}_t \boldsymbol{\omega}_t^T = V^\omega + S_t^\omega \quad \forall t \geq 1 \quad (12)$$

where V^ω and S_t^ω are random matrices of the same dimension as $\boldsymbol{\omega}_t \boldsymbol{\omega}_t^T$, $E(\mathbf{vec}(S_t^\omega)) = 0$ and $\text{Cov}(\mathbf{vec}(V^\omega), \mathbf{vec}(S_t^\omega)) = \text{Cov}(\mathbf{vec}(S_t^\omega), \mathbf{vec}(S_s^\omega)) = 0$, $\forall s \neq t$, $\text{Var}(\mathbf{vec} S_s^\omega) = \text{Var}(\mathbf{vec} S_t^\omega)$, $\forall s, t$. Decomposing $\mathbf{vec}(\boldsymbol{\nu}_t \boldsymbol{\nu}_t^T)$ similarly, we obtain

$$\boldsymbol{\nu}_t \boldsymbol{\nu}_t^T = V^\nu + S_t^\nu \quad \forall t \geq 1 \quad (13)$$

with properties as for representation (12). Note that $E(V^\omega) = E(\boldsymbol{\omega}_t \boldsymbol{\omega}_t^T) = \text{Var}(\boldsymbol{\omega}_t) = W$ and so learning about V^ω will allow us to learn about the covariance matrix for the residuals for the state, and $E(V^\nu) = E(\boldsymbol{\nu}_t \boldsymbol{\nu}_t^T) = \text{Var}(\boldsymbol{\nu}_t) = V$, and so learning about V^ν will allow us to learn about the covariance matrix for the observational residuals. Representations (12) and (13) decompose our uncertainty about $\boldsymbol{\omega}_t \boldsymbol{\omega}_t^T$ and $\boldsymbol{\nu}_t \boldsymbol{\nu}_t^T$ into two parts. Bayes linear updating (with enough data) will eliminate the aspects of uncertainty derived from uncertainty about V^ω and V^ν .

In order to conduct a Bayes linear analysis on the quadratic structure we need additional covariance specifications $\text{Var}(\mathbf{vec} V^\omega)$, $\text{Var}(\mathbf{vec} V^\nu)$, $\text{Var}(\mathbf{vec} S_t^\omega)$ and $\text{Var}(\mathbf{vec} S_t^\nu)$, for some t .

3.2 Example

In our example the \mathbf{X}_t vector is 6-dimensional, and so the matrices, V^ω , V^ν , S^ω and S^ν are 6×6 -dimensional. Consequently, the matrices $\text{Var}(\mathbf{vec} V^\omega)$, $\text{Var}(\mathbf{vec} V^\nu)$, $\text{Var}(\mathbf{vec} S_t^\omega)$ and $\text{Var}(\mathbf{vec} S_t^\nu)$ are 36×36 -dimensional. When referring to the components of $\text{Var}(\mathbf{vec} V^\omega)$, the notation v_{ijkl}^ω will be used to denote the covariance between the $(i, j)^{th}$ and $(k, l)^{th}$ elements of V^ω . Similar notation is used for $\text{Var}(\mathbf{vec} V^\nu)$. Also s_{ijkl}^ω and s_{ijkl}^ν are used for the components of $\text{Var}(\mathbf{vec} S_t^\omega)$ and $\text{Var}(\mathbf{vec} S_t^\nu)$ respectively. The following covariance specifications were made for our example:

$$v_{iiii}^\omega = 9/4, \forall i, \quad v_{ijij}^\omega = 9/16, \forall i \neq j, \quad v_{iijj}^\omega = 1/5, \forall i \neq j, \quad (14)$$

$$v_{iiii}^\nu = 25, \forall i, \quad v_{ijij}^\nu = 1, \forall i \neq j, \quad v_{iijj}^\nu = 4, \forall i \neq j, \quad (15)$$

$$s_{iiii}^\omega = 30, \forall i, \quad s_{ijij}^\omega = 15, \forall i \neq j, \quad s_{iijj}^\omega = 2500, \forall i, \quad s_{ijij}^\nu = 1000, \forall i \neq j. \quad (16)$$

For instance, v_{iiii}^ω is the variance specification for the $(i, i)^{th}$ element of V^ω , which represents the underlying variance of the i^{th} element of $\boldsymbol{\omega}_t$. From (9), it has expectation 4. From (6), this value governs the rate of change of $\boldsymbol{\Theta}_t$. By considering the range of plausible variances for the way $\boldsymbol{\Theta}_t$ might change over time, it was felt reasonable that a standard deviation specification of $3/2$ should be made. The other specifications were made in a similar fashion. For simplicity in this example, the specifications for s_{ijkl}^ω and s_{ijkl}^ν were made

using the fourth moments of the multivariate normal distribution compatible with the given second-order structure as a guide.

While we would require considerably more specifications for a full Bayes or Bayes linear analysis, (14), (15), and (16) are sufficient for our purposes as these are the only specifications needed for the matrix object approach to belief revision which we are to take in the later sections.

There is a lot of symmetry in these values, greatly simplifying the specification, but once again, any non-negative covariance structure over the quadratic products is acceptable. Here we have used assumptions of exchangeability over the variances and the covariances. Many of the specifications made for the quadratic structure will be “averaged over” in the matrix object approach to covariance adjustment which we shall develop, and so there is a limit to the effort that we may wish to put into very detailed specifications at this stage since our suggested analysis will not be overly sensitive to the individual specifications.

3.3 Observable quadratic terms

We will first construct certain linear combinations of the observables which do not involve the state vector, Θ_t . This is useful for various reasons, and in particular because it greatly reduces the prior specification required for the analysis of the quadratic structure. In this paper, we shall mainly be concerned with DLMs for which there exists an $r \times r$ matrix H , such that $HF^T = F^TG$. We call such DLMs *two-step invertible*. Note that a DLM will be two-step invertible if F is of full rank and $r \geq p$ (as will often be the case for high-dimensional time series), and there will often be many matrices H satisfying $HF^T = F^TG$. For example, $H = F^TG^TF^{T\dagger}$ (where $F^{T\dagger}$ represents any generalised inverse of F^T) is a solution. Further, if F is of full rank, $r < p$ and such a matrix exists, then $H = F^TGF(F^TF)^{-1}$, and so H exists, if and only if $F^TGF(F^TF)^{-1}F^T = F^TG$. Note also that the matrix H^2 has the property that $H^2F^T = F^TG^2$. For a two-step invertible DLM, we may construct the following vectors of observables which do not involve the state vector:

$$\mathbf{X}'_t = \mathbf{X}_t - H\mathbf{X}_{t-1} = F^T\boldsymbol{\omega}_t + \boldsymbol{\nu}_t - H\boldsymbol{\nu}_{t-1} \quad \forall t \geq 2 \quad (17)$$

$$\mathbf{X}''_t = \mathbf{X}_t - H^2\mathbf{X}_{t-2} = F^T\boldsymbol{\omega}_t + F^TG\boldsymbol{\omega}_{t-1} + \boldsymbol{\nu}_t - H^2\boldsymbol{\nu}_{t-2} \quad \forall t \geq 3 \quad (18)$$

We form the matrices of quadratic products, $\mathbf{X}'_t\mathbf{X}'_t{}^T$ and $\mathbf{X}''_t\mathbf{X}''_t{}^T \quad \forall t$. As we shall see, these are predictive for V^ω and V^ν .

Not all DLMs are two-step invertible, but for the constant dynamic linear model outlined in Section 2.1, it is always possible to construct linear combinations of the observations which do not involve the state, provided only that the constant dynamic linear model in question is *observable* (for a discussion of the very weak restriction of observability, see West and Harrison (1989), Chapter 5). However, in general such linear combinations require more than two successive observations from the series. Consequently, for simplicity

we restrict attention to the two-step invertible model. However, the approach is quite general, and may be applied similarly for any constant observable DLM, the only difference being that the covariance specification is more complicated, and more quantities are involved in the adjustment.

3.4 Example

For our example, F , G and H are all the identity, and so we form the one and two-step differences of the observables:

$$\mathbf{X}_t^{(1)} = \mathbf{X}_t - \mathbf{X}_{t-1} = \boldsymbol{\omega}_t + \boldsymbol{\nu}_t - \boldsymbol{\nu}_{t-1} \quad \forall t \geq 2 \quad (19)$$

$$\mathbf{X}_t^{(2)} = \mathbf{X}_t - \mathbf{X}_{t-2} = \boldsymbol{\omega}_t + \boldsymbol{\omega}_{t-1} + \boldsymbol{\nu}_t - \boldsymbol{\nu}_{t-2} \quad \forall t \geq 3 \quad (20)$$

We then form the quadratic products of these, $\mathbf{X}_t^{(1)} \mathbf{X}_t^{(1)\top}$ and $\mathbf{X}_t^{(2)} \mathbf{X}_t^{(2)\top}$. These observables are predictive for V^ω and V^ν and so may be used to learn about the underlying covariance structure. All means and covariances that we require for the subsequent analysis are determined by specifications in (7), (8), (9), (10), (14), (15) and (16). The precise form of the covariance structure over the observables is rather complex, and is given in the appendix.

4 n -step exchangeability

4.1 n -step exchangeable collections

The covariance structure over \mathbf{X}'_t , \mathbf{X}''_t and their quadratic products has a general structure common amongst ordered vectors of random quantities of the type arising from differenced time series. The covariance structure is invariant under arbitrary translations and reflections of the ordering, and the auto-correlation function becomes constant after some distance, n . We call ordered vectors with this property, *second-order n -step exchangeable*. Covariance may be interpreted as an inner-product on a space of random quantities. We will find it useful to consider inner-products on spaces of more general random entities. In particular, in Section 5.1 we shall define an inner-product on a space of random matrices. We require a concept of n -step exchangeability which is sufficiently general that it is also valid for this space of matrices, and so we formalise the concept as follows.

Let $\{Y_{jk} | \forall j, k\}$ be a collection of random entities of interest to us. Also form a maximal linearly independent collection of constant entities of the same type, and call this collection $C = [C_1, C_2, \dots]$. When we are dealing with random scalars, C will consist of the single scalar, $C_1 = 1$. In Section 5.1, we describe the constant space for a collection of random matrices. Form the vector space

$$\mathcal{V} = \text{span}\{C_j, Y_{jk} | \forall j, k\} \quad (21)$$

so that the random entities are now vectors within this space. Define an inner-product $(\cdot, \cdot) : \mathcal{V} \times \mathcal{V} \longrightarrow \mathbf{R}$ on \mathcal{V} . The inner-product should capture certain aspects of our beliefs about the relationships between the elements of \mathcal{V} ; for scalars, we might use $(X, Y) = E(XY)$. Form the completion of the space \mathcal{V} , and denote this Hilbert space by \mathcal{H} . Also define a bounded linear expectation function on the elements of \mathcal{V} . For the purposes of this paper it suffices to think in terms of the usual definition of expectation. In general, however, we define a bounded linear function $E(\cdot) : \mathcal{H} \longrightarrow \text{span}\{C\}$, such that $\forall Y \in \mathcal{H}$, $E(Y)$ is the orthogonal projection of Y into $\text{span}\{C\}$, with respect to the inner-product (\cdot, \cdot) . This is the generalisation of expectation which we require for general spaces, and is defined in terms of the inner-product. It coincides with the usual definitions for all of the examples in this paper.

If $\exists n \in \mathbf{N}$ such that $E(Y_{jk}) = e_j \quad \forall j, k$, and

$$\begin{aligned} (Y_{ik}, Y_{jl}) &= d_{0ij} & \forall i, j, |k-l| = 0 \\ (Y_{ik}, Y_{jl}) &= d_{1ij} & \forall i, j, |k-l| = 1 \\ &\vdots & \vdots \\ (Y_{ik}, Y_{jl}) &= d_{(n-1)ij} & \forall i, j, |k-l| = n-1 \\ (Y_{ik}, Y_{jl}) &= c_{ij} & \forall i, j, |k-l| \geq n \end{aligned} \tag{22}$$

the collection $\{Y_{jk} | \forall j, k\}$ is said to be *generalised second-order n -step exchangeable over k* .

We apply this definition to the elements, $X^* = \{X'_{it}X'_{jt} | \forall i, j, \forall t \geq 2\}$ and $X^{**} = \{X''_{it}X''_{jt} | \forall i, j, \forall t \geq 3\}$, of the matrices $\{\mathbf{X}'_t \mathbf{X}'_t{}^T | \forall t \geq 2\}$ and $\{\mathbf{X}''_t \mathbf{X}''_t{}^T | \forall t \geq 3\}$ defined in Section 3.3, where X'_{it} denotes the i^{th} component of the vector \mathbf{X}'_t . Form a vector space, \mathcal{V} consisting of all linear combinations of the elements of X^* and X^{**} and the unit constant, and define the inner-product on this space as $(X, Y) = E(XY)$, $\forall X, Y \in \mathcal{V}$. We may easily check that X^* is (second order) 2-step exchangeable over t , and that X^{**} is 3-step exchangeable over t .

4.2 Representation for n -step exchangeable collections

Goldstein (1986) constructs a general representation for second-order exchangeable collections. There is an analogous representation for collections with the weaker property of n -step exchangeability, constructed in a similar way.

Theorem 1 *Let $\{Y_{jk} | \forall j, k\}$ be generalised second-order n -step exchangeable over k with respect to the inner-product (\cdot, \cdot) . Then the Y_{jk} may be represented as*

$$Y_{jk} = M_j + R_{jk} \quad \forall j, k \tag{23}$$

where the M_j and R_{jk} have the following properties:

$$E(Y_{jk}) = E(M_j), E(R_{jk}) = 0 \quad \forall j, k \tag{24}$$

$$(M_i, M_j) = (M_i, Y_{jk}) = c_{ij}, \quad (M_i, R_{jk}) = 0 \quad \forall i, j, k \quad (25)$$

$$(R_{ik}, R_{jl}) = (Y_{ik}, R_{jl}) = (Y_{ik}, Y_{jl}) - c_{ij} \quad \forall i, j, k, l \quad (26)$$

Further, the $\{R_{jk} | \forall j, k\}$ are generalised second-order n -step exchangeable over k , with $(R_{ik}, R_{jl}) = 0 \quad \forall i, j, |k-l| \geq n$.

Proof

Let

$$M_{im} = \frac{1}{m} \sum_{k=1}^m Y_{ik} \quad \forall i, m \quad (27)$$

Observe that the sequence M_{i1}, M_{i2}, \dots is *Cauchy* $\forall i$ ie. that $(M_{ik} - M_{il}, M_{ik} - M_{il}) \longrightarrow 0$ as $k, l \longrightarrow \infty$, which follows directly from the properties of n -step exchangeable sequences. Construct the quantity M_i to be the Cauchy limit of this sequence so that

$$\lim_{m \rightarrow \infty} (M_{im}, Y) = (M_i, Y) \quad \forall i, \forall Y \in \mathcal{H} \quad (28)$$

Linearity of $E(\cdot)$ gives $E(M_{im}) = e_i \quad \forall i, m$, and hence applying (28) for $Y \in C$ we deduce $E(M_i) = e_i$. Define R_{ik} via $R_{ik} = X_{ik} - M_i \quad \forall i$, so that $E(R_{im}) = 0 \quad \forall i, m$. The other properties of the representation follow directly from (28). \square

As for the case of second-order exchangeability, the mean components of the representation, M_j , represent the quantities which we may learn about by linear fitting on the data. We may resolve as much uncertainty as we wish about these quantities given a sufficient number of observations, by such linear fitting. We therefore say that the n -step exchangeable collection $\{Y_{jk} | \forall j, k\}$ with representation $Y_{jk} = M_j + R_{jk} \quad \forall j, k$ identify the random quantities $M_j, \forall j$.

4.3 Identification of the covariance structure underlying the DLM

The n -step exchangeability representation theorem allows us to construct models for the observable quadratic products we have formed. The elements of the collection $\{\mathbf{X}'_t \mathbf{X}'^T_t | \forall t \geq 2\}$ for the two-step invertible DLM, are 2-step exchangeable over t . From Theorem 1, we may construct the representation (23). The identified quantities may be constructed as the Cauchy limit of the arithmetic means of the elements.

Lemma 2 *The 2-step exchangeable collection $\{\mathbf{X}'_t \mathbf{X}'^T_t | \forall t \geq 2\}$ identify the matrix*

$$M' = F^T V^\omega F + V^\nu + H V^\nu H^T \quad (29)$$

and the 3-step exchangeable collection $\{\mathbf{X}''_t \mathbf{X}''^T_t | \forall t \geq 3\}$ identify

$$M'' = F^T G V^\omega G^T F + F^T V^\omega F + V^\nu + H^2 V^\nu H^{2T} \quad (30)$$

Proof

$$M' = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=2}^N \mathbf{X}'_t \mathbf{X}'_t{}^T \quad (31)$$

$$= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=2}^N (F^T \boldsymbol{\omega}_t + \boldsymbol{\nu}_t - H \boldsymbol{\nu}_{t-1}) (F^T \boldsymbol{\omega}_t + \boldsymbol{\nu}_t - H \boldsymbol{\nu}_{t-1})^T \quad (32)$$

$$= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=2}^N F^T \boldsymbol{\omega}_t \boldsymbol{\omega}_t^T F + \boldsymbol{\nu}_t \boldsymbol{\nu}_t^T + H \boldsymbol{\nu}_{t-1} \boldsymbol{\nu}_{t-1}^T H^T \quad (33)$$

$$= F^T V^\omega F + V^\nu + H V^\nu H^T \quad (34)$$

The derivation of M'' is similar. \square

Now since $\frac{1}{2} [H M' H^T - (M'' - M')] = H V^\nu H^T$ we deduce that for a 2-step invertible series, the collection

$$\left\{ \frac{1}{2} [H \mathbf{X}'_t \mathbf{X}'_t{}^T H^T - (\mathbf{X}''_t \mathbf{X}''_t{}^T - \mathbf{X}'_t \mathbf{X}'_t{}^T)] \mid \forall t \geq 3 \right\} \quad (35)$$

identifies $H V^\nu H^T$.

Now if $r \geq p$, H can usually be chosen to be invertible. If $r < p$, since we assume that F and G are of full rank, so is H , and so H is invertible. In this paper we restrict attention to those two-step invertible DLMS which have an invertible H . Consequently, since $\frac{1}{2}[M' - H^{-1}(M'' - M')H^{-1T}] = V^\nu$ and $M' - V^\nu - H V^\nu H^T = F^T V^\omega F$ we get

Theorem 3 *For a 2-step invertible series with invertible H , put*

$$M_t = \frac{1}{2} [\mathbf{X}'_t \mathbf{X}'_t{}^T - H^{-1}(\mathbf{X}''_t \mathbf{X}''_t{}^T - \mathbf{X}'_t \mathbf{X}'_t{}^T)H^{-1T}] \quad (36)$$

(i) *The collection of matrices $\{M_t \mid \forall t \geq 3\}$ identify V^ν .*

(ii) *The collection of matrices $\{\mathbf{X}'_t \mathbf{X}'_t{}^T - M_t - H M_t H^T \mid \forall t \geq 3\}$ identify $F^T V^\omega F$.*

\square

If $p > r$, we do not identify the entire matrix V^ω . This would require a fuller selection of observables than we have considered here. The identified matrix, $F^T V^\omega F$, is the contribution to the uncertainty for \mathbf{X}_t from $\boldsymbol{\Theta}_t$, given $\boldsymbol{\Theta}_{t-1}$.

4.4 Example

In our example, Theorem 3 implies that the collection $\{\mathbf{X}_t^{(2)} \mathbf{X}_t^{(2)T} - \mathbf{X}_t^{(1)} \mathbf{X}_t^{(1)T} \mid \forall t \geq 3\}$ identify V^ω and that the collection $\{\mathbf{X}_t^{(1)} \mathbf{X}_t^{(1)T} - \frac{1}{2} \mathbf{X}_t^{(2)} \mathbf{X}_t^{(2)T} \mid \forall t \geq 3\}$ identify V^ν .

By observing sales at an increasing (but finite) number of time points, we may resolve through linear fitting, as much uncertainty as we wish about the underlying covariance structure for the particular time series model we are dealing with.

If all fourth order prior belief specifications have been made, a simple Bayes linear analysis can be carried out in order to learn about the underlying covariance structure by adjusting the elements of V^ν , V^ω by the elements of the observable matrices $\mathbf{X}_t^{(1)}\mathbf{X}_t^{(1)\top}$, $\mathbf{X}_t^{(2)}\mathbf{X}_t^{(2)\top}$. However, for long, high-dimensional time series, the number of quantities involved in a full linear adjustment is extremely large, and so it is important to reduce the dimensionality of the problem, and preserve the inherent matrix structure.

5 Matrix objects for the time series

5.1 Formation of the matrix space

We are interested in learning about $r \times r$ dimensional covariance matrices. We first form a collection of r^2 linearly independent $r \times r$ constant matrices such that $C_{r(i-1)+j}$ is the matrix with a 1 in the $(i, j)^{th}$ position, and zeros elsewhere, where i and j range from 1 to r and call this collection $C = [C_1, \dots, C_{r^2}]$. This collection of matrices is a basis for the space of *known* $r \times r$ matrices. Define the collections of matrices

$$X_2^\dagger = \{\mathbf{X}_2'\mathbf{X}_2'^\top, H\mathbf{X}_2'\mathbf{X}_2'^\top H^\top, H^{-1}\mathbf{X}_2'\mathbf{X}_2'^\top H^{-1\top}\} \quad (37)$$

$$X_t^\dagger = \{\mathbf{X}_t'\mathbf{X}_t'^\top, \mathbf{X}_t''\mathbf{X}_t''^\top, H\mathbf{X}_t'\mathbf{X}_t'^\top H^\top, H^{-1}\mathbf{X}_t'\mathbf{X}_t'^\top H^{-1\top}, H^{-1}\mathbf{X}_t''\mathbf{X}_t''^\top H^{-1\top}\}, \quad \forall t \geq 3 \quad (38)$$

Form the real vector space, \mathcal{N} whose elements are linear combinations of random $r \times r$ matrices as follows.

$$\mathcal{N} = \text{span} \left\{ C, X_2^\dagger, X_3^\dagger, \dots \right\} \quad (39)$$

Define an inner-product on \mathcal{N} via

$$(A, B) = E(\text{Tr}[AB^\top]), \quad \forall A, B \in \mathcal{N} \quad (40)$$

This inner-product is discussed and motivated in Wilkinson and Goldstein (1995). Complete \mathcal{N} into a Hilbert space, \mathcal{M} . When the space is completed, limit points such as $HV^\nu H^\top$, V^ν , and $F^\top V^\omega F$ are added to the space. The inner-product on this space is determined by our beliefs about the quadratic products, since

$$(A, B) = \sum_{j=1}^r \sum_{k=1}^r [\text{Cov}(A_{jk}, B_{jk}) + E(A_{jk})E(B_{jk})] \quad \forall A, B \in \mathcal{M} \quad (41)$$

We may carry out Bayes linear adjustment in this space by orthogonal projection of the matrices of interest into subspaces of observable matrices. Our adjusted expectations for these matrices are linear combinations of the prior matrices and the observable matrices. Note that this matrix approach to belief adjustment is a more direct way of getting at desirable linearity properties of conditional expectations for matrices, than via somewhat artificial constructs such as the matrix normal distribution.

5.2 Example

For our example, we simply construct

$$\mathcal{N} = \text{span}\{C, \mathbf{X}_2^{(1)} \mathbf{X}_2^{(1)\text{T}}, \mathbf{X}_3^{(1)} \mathbf{X}_3^{(1)\text{T}}, \dots, \mathbf{X}_3^{(2)} \mathbf{X}_3^{(2)\text{T}}, \mathbf{X}_4^{(2)} \mathbf{X}_4^{(2)\text{T}}, \dots\} \quad (42)$$

and impose the inner-product (40), inducing the Hilbert space \mathcal{M} , which contains limit points such as V^ν and V^ω . Note that in order to evaluate (40), the specifications needed are precisely those which were made in Section 3.2. The fact that many other aspects of the fourth order specifications are not necessary is very helpful, as this greatly reduces the specification burden. Often it is most straightforward to make direct primitive specifications for the matrix object inner-product. However, for simplicity in this paper, we have built up the specifications for the matrix inner product from specifications over the scalar quadratic products, thus establishing the links between the scalar and matrix analysis. This is analogous to specifying an expectation of a random quantity by breaking it up over a partition of events and specifying probabilities over the partition.

5.3 n -step exchangeable matrix objects

The definition of generalised n -step exchangeability applies directly to matrix objects in the space \mathcal{M} . The collection of matrix objects $\{\mathbf{X}'_t \mathbf{X}'_t{}^{\text{T}} | \forall t \geq 2\}$ is 2-step exchangeable in the space \mathcal{M} , and the collection $\{\mathbf{X}''_t \mathbf{X}''_t{}^{\text{T}} | \forall t \geq 3\}$ is 3-step exchangeable. This leads to a restatement of Theorem 3 for matrices in the space \mathcal{M} . The limit points are the matrices of limit points of their elements, due to the consistency of the inner-products on the scalar and matrix spaces.

Theorem 4 Put $M_t = \frac{1}{2} [\mathbf{X}'_t \mathbf{X}'_t{}^{\text{T}} - H^{-1}(\mathbf{X}''_t \mathbf{X}''_t{}^{\text{T}} - \mathbf{X}'_t \mathbf{X}'_t{}^{\text{T}})H^{-1\text{T}}]$.

- (i) The collection $\{M_t | \forall t \geq 3\}$ identifies V^ν in \mathcal{M} .
- (ii) The collection $\{\mathbf{X}'_t \mathbf{X}'_t{}^{\text{T}} - M_t - H M_t H^{\text{T}} | \forall t \geq 3\}$ identifies $F^{\text{T}} V^\omega F$ in \mathcal{M} .

□

Now, for any subspace D such that $C \subseteq D \subseteq \mathcal{M}$, we define *adjusted* matrix expectation, $E_D : \mathcal{M} \rightarrow D$ to be such that $\forall Y \in \mathcal{M}$, $E_D(Y)$ is the orthogonal projection of Y into D , with respect to the inner-product (\cdot, \cdot) . If $D = \text{span}\{C\}$ then we write $E(\cdot)$ for $E_D(\cdot)$, as this is the usual matrix expectation, consisting of the matrix of expectations of elements.

5.4 Adjustment

Suppose that we are considering observing $n > 3$ time points in the series. Form the matrix space, \mathcal{M} , and the observable subspace $D_n \subseteq \mathcal{M}$

$$D_n = \text{span}\{C, X_2^\dagger, X_3^\dagger, \dots, X_n^\dagger\} \quad (43)$$

Then the adjusted expectation map, $E_{D_n}(\cdot) : \mathcal{M} \rightarrow D_n$, is the orthogonal projection into the D_n space. In particular, we evaluate $E_{D_n}(V^\nu)$, $E_{D_n}(F^T V^\omega F)$, which are matrices in the D_n space. These adjusted expectations are analogous to posterior expected values for the matrix objects, taking into account only those limited aspects of the problem which we have considered. The general relationships between adjusted and posterior beliefs is discussed in Goldstein (1994).

6 Bayes linear adjustment for the example

6.1 The adjusted covariance matrices

Adjustments were carried out using 17 time points from the actual time series. The matrix objects V^ω and V^ν were adjusted in the following ways:

$$E_D(V^\omega) = \begin{pmatrix} 4.8 & 0.9 & 1.0 & 1.0 & 0.8 & 1.5 \\ 0.9 & 3.9 & 1.2 & 0.9 & 1.1 & 0.3 \\ 1.0 & 1.2 & 4.0 & 1.1 & 1.1 & 0.7 \\ 1.0 & 0.9 & 1.1 & 6.8 & 0.7 & 0.8 \\ 0.8 & 1.1 & 1.1 & 0.7 & 3.9 & 0.8 \\ 1.5 & 0.3 & 0.7 & 0.8 & 0.8 & 4.7 \end{pmatrix} \quad (44)$$

$$E_D(V^\nu) = \begin{pmatrix} 41.8 & -5.4 & -4.4 & -8.0 & -4.7 & -2.4 \\ -5.4 & 36.7 & -3.8 & -0.2 & -3.2 & -4.1 \\ -4.4 & -3.8 & 36.1 & -4.4 & -3.5 & -7.5 \\ -8.0 & -0.2 & -4.4 & 56.6 & -5.6 & 4.8 \\ -4.7 & -3.2 & -3.5 & -5.6 & 34.9 & -4.9 \\ -2.4 & -4.1 & -7.5 & 4.8 & -4.9 & 44.0 \end{pmatrix} \quad (45)$$

Comparing these with their prior specifications, given in (9) and (10), we see that the adjusted matrices are perturbations of the prior expectations for the matrices. Notice that the variance associated with the fourth variable has been inflated considerably in both matrices. The sample variances for the 17 cases of the six brands we considered were 167, 22, 37, 560, 18 and 427. Informally, it seems that there may indeed be more variability associated with the fourth (and sixth) variable.

6.2 First order adjustment

Since our aim is to predict sales more accurately, a sensible test of the procedure is to compare the performance of the first order model, (5), (6), using both the prior and adjusted covariance matrices. We find that

the Bayes linear diagnostic warnings (the *size* and *bearings* of the adjustments, as described in Goldstein (1988)) are noticeably closer to their expected values when using the adjusted matrices. For the given example, most of the *size ratios* for adjustments of the first order structure were noticeably closer to one using the adjusted covariance structure to predict future values, suggesting that the adjusted matrices match more closely with the forecast performance of the model.

7 Conclusions

Good forecasting requires careful updating of the covariances within the time series structure. Informally, the degree of *shrinkage* between the prior and the data is updated, and relationships between variables are properly taken into account. Since we are able to adjust the covariance matrices for both the *observational* as well as *state* residuals, we are able to properly understand the *competition* and *demand* effects taking place within the series. By taking a matrix object approach, we greatly simplify the problem by reducing dimensionality. This is important for both simplifying belief specification and belief adjustment, and also for interpretation of the structure of the adjustment and accompanying diagnostics. There are also the general advantages of the Bayes linear approach; namely of allowing complete flexibility for the prior specifications, without placing distributional restrictions on the data or model components.

8 Acknowledgements

The first named author is supported by a grant from the UK's EPSRC. All Bayes linear computations were carried out using the Bayes linear computing package, [B/D] described in Wooff (1992), and explained in detail in Goldstein and Wooff (1995). The covariance calculations for the quadratic products were carried out using the REDUCE computer algebra system. The example data was supplied by *Positive Concepts Ltd.*

Appendix

Covariance structure for the example

We give here the full covariance structure over the quadratic products of the one and two step differences, $\mathbf{X}_t^{(1)}$, $\mathbf{X}_t^{(2)}$ defined by (19), (20). First we need some notation for two different “direct products” of matrices.

Definition 5 For $r \times r$ matrices A (having entry a_{ij} in row i , column j) and B (having entry b_{ij} in row i , column j) we define the (left) tensor product, $A \otimes B$ of A and B to be the $r^2 \times r^2$ matrix with the element $a_{jk}b_{lm}$ in row $r(l-1) + j$, column $r(m-1) + k$.

Definition 6 For $r \times r$ matrices A (having entry a_{ij} in row i , column j) and B (having entry b_{ij} in row i , column j) we define the star product $A \star B$ of A and B to be the $r^2 \times r^2$ matrix with the element $a_{jk}b_{lm}$ in row $r(l-1)+j$, column $r(k-1)+m$.

Given these definitions, the covariance structure over the quadratic products of the 1-step differences is determined by the following relations:

$$\text{Cov}(\text{vec}V^\omega, \text{vec}(\mathbf{X}_t^{(1)} \mathbf{X}_t^{(1)\text{T}})) = \text{Var}(\text{vec}V^\omega) \quad (46)$$

$$\text{Cov}(\text{vec}V^\nu, \text{vec}(\mathbf{X}_t^{(1)} \mathbf{X}_t^{(1)\text{T}})) = 2\text{Var}(\text{vec}V^\nu) \quad (47)$$

$$\begin{aligned} \text{Cov}(\text{vec}(\mathbf{X}_t^{(1)} \mathbf{X}_t^{(1)\text{T}}), \text{vec}(\mathbf{X}_t^{(1)} \mathbf{X}_t^{(1)\text{T}})) &= \text{Var}(\text{vec}V^\omega) + 4\text{Var}(\text{vec}V^\nu) \\ &+ \text{Var}(\text{vec}S_t^\nu) + \text{Var}(\text{vec}S_{t-1}^\nu) \\ &+ \text{Var}(\text{vec}S_t^\omega) + 2[\text{E}(V^\nu \otimes V^\nu) + \text{E}(V^\nu \star V^\nu)] \\ &+ 4[\text{E}(V^\nu) \otimes \text{E}(V^\omega) + \text{E}(V^\nu) \star \text{E}(V^\omega)] \\ &+ \text{E}(V^\omega) \otimes \text{E}(V^\nu) + \text{E}(V^\omega) \star \text{E}(V^\nu) \end{aligned} \quad (48)$$

$$\text{Cov}(\text{vec}(\mathbf{X}_t^{(1)} \mathbf{X}_t^{(1)\text{T}}), \text{vec}(\mathbf{X}_{t-1}^{(1)} \mathbf{X}_{t-1}^{(1)\text{T}})) = 4(\text{Var}(\text{vec}V^\nu) + \text{Var}(\text{vec}V^\omega)) + \text{Var}(\text{vec}S_{t-1}^\nu) \quad (49)$$

$$\text{Cov}(\text{vec}(\mathbf{X}_t^{(1)} \mathbf{X}_t^{(1)\text{T}}), \text{vec}(\mathbf{X}_{t-s}^{(1)} \mathbf{X}_{t-s}^{(1)\text{T}})) = 4\text{Var}(\text{vec}V^\nu) + \text{Var}(\text{vec}V^\omega) \quad \forall t, \forall s \geq 2 \quad (50)$$

The covariance structure over the quadratic products of the 2-step differences are given below.

$$\text{Cov}(\text{vec}V^\omega, \text{vec}(\mathbf{X}_t^{(2)} \mathbf{X}_t^{(2)\text{T}})) = 2\text{Var}(\text{vec}V^\omega) \quad (51)$$

$$\text{Cov}(\text{vec}V^\nu, \text{vec}(\mathbf{X}_t^{(2)} \mathbf{X}_t^{(2)\text{T}})) = 2\text{Var}(\text{vec}V^\nu) \quad (52)$$

$$\begin{aligned} \text{Cov}(\text{vec}(\mathbf{X}_t^{(2)} \mathbf{X}_t^{(2)\text{T}}), \text{vec}(\mathbf{X}_t^{(2)} \mathbf{X}_t^{(2)\text{T}})) &= 4\text{Var}(\text{vec}V^\omega) + 4\text{Var}(\text{vec}V^\nu) + \text{Var}(\text{vec}S_t^\nu) \\ &+ \text{Var}(\text{vec}S_{t-2}^\nu) + \text{Var}(\text{vec}S_t^\omega) + \text{Var}(\text{vec}S_{t-1}^\omega) \\ &+ 2[\text{E}(V^\nu \otimes V^\nu) + \text{E}(V^\nu \star V^\nu) + \text{E}(V^\omega \otimes V^\omega) \\ &+ \text{E}(V^\omega \star V^\omega)] + 4[\text{E}(V^\nu) \otimes \text{E}(V^\omega) + \text{E}(V^\nu) \star \text{E}(V^\omega)] \\ &+ \text{E}(V^\omega) \otimes \text{E}(V^\nu) + \text{E}(V^\omega) \star \text{E}(V^\nu) \end{aligned} \quad (53)$$

$$\text{Cov}(\text{vec}(\mathbf{X}_t^{(2)} \mathbf{X}_t^{(2)\text{T}}), \text{vec}(\mathbf{X}_{t-1}^{(2)} \mathbf{X}_{t-1}^{(2)\text{T}})) = 4[\text{Var}(\text{vec}V^\nu) + \text{Var}(\text{vec}V^\omega)] + \text{Var}(\text{vec}S_{t-1}^\omega) \quad (54)$$

$$\text{Cov}(\text{vec}(\mathbf{X}_t^{(2)} \mathbf{X}_t^{(2)\text{T}}), \text{vec}(\mathbf{X}_{t-2}^{(2)} \mathbf{X}_{t-2}^{(2)\text{T}})) = 4[\text{Var}(\text{vec}V^\nu) + \text{Var}(\text{vec}V^\omega)] + \text{Var}(\text{vec}S_{t-2}^\nu) \quad (55)$$

$$\text{Cov}(\text{vec}(\mathbf{X}_t^{(2)} \mathbf{X}_t^{(2)\text{T}}), \text{vec}(\mathbf{X}_{t-s}^{(2)} \mathbf{X}_{t-s}^{(2)\text{T}})) = 4\text{Var}(\text{vec}V^\nu) + \text{Var}(\text{vec}V^\omega) \quad \forall t, \forall s \geq 3 \quad (56)$$

The covariances between the one and two step differences are determined as follows:

$$\text{Cov}(\text{vec}(\mathbf{X}_t^{(1)} \mathbf{X}_t^{(1)\text{T}}), \text{vec}(\mathbf{X}_{t+s}^{(2)} \mathbf{X}_{t+s}^{(2)\text{T}})) = 4\text{Var}(\text{vec}V^\nu) + 2\text{Var}(\text{vec}V^\omega) \quad \forall t, \forall s \geq 3 \quad (57)$$

$$\text{Cov}(\text{vec}(\mathbf{X}_t^{(1)} \mathbf{X}_t^{(1)\text{T}}), \text{vec}(\mathbf{X}_{t+2}^{(2)} \mathbf{X}_{t+2}^{(2)\text{T}})) = 4\text{Var}(\text{vec}V^\nu) + 2\text{Var}(\text{vec}V^\omega) + \text{Var}(\text{vec}S_{t-2}^\omega) \quad (58)$$

$$\begin{aligned} \text{Cov}(\text{vec}(\mathbf{X}_t^{(1)} \mathbf{X}_t^{(1)\text{T}}), \text{vec}(\mathbf{X}_{t+1}^{(2)} \mathbf{X}_{t+1}^{(2)\text{T}})) &= 2\text{Var}(\text{vec}V^\omega) + 4\text{Var}(\text{vec}V^\nu) \\ &+ \text{Var}(\text{vec}S_{t-2}^\nu) + \text{Var}(\text{vec}S_{t-1}^\omega) \\ &+ \text{E}(V^\nu) \otimes \text{E}(V^\omega) + \text{E}(V^\nu) \star \text{E}(V^\omega) \\ &+ \text{E}(V^\omega) \otimes \text{E}(V^\nu) + \text{E}(V^\omega) \star \text{E}(V^\nu) \end{aligned} \quad (59)$$

$$\begin{aligned}
\text{Cov}(\text{vec}(\mathbf{X}_t^{(1)} \mathbf{X}_t^{(1)\text{T}}), \text{vec}(\mathbf{X}_t^{(2)} \mathbf{X}_t^{(2)\text{T}})) &= 2\text{Var}(\text{vec}V^\omega) + 4\text{Var}(\text{vec}V^\nu) \\
&+ \text{Var}(\text{vec}S_t^\nu) + \text{Var}(\text{vec}S_t^\omega) \\
&+ \text{E}(V^\nu) \otimes \text{E}(V^\omega) + \text{E}(V^\nu) \star \text{E}(V^\omega) \\
&+ \text{E}(V^\omega) \otimes \text{E}(V^\nu) + \text{E}(V^\omega) \star \text{E}(V^\nu)
\end{aligned} \tag{60}$$

$$\text{Cov}(\text{vec}(\mathbf{X}_t^{(1)} \mathbf{X}_t^{(1)\text{T}}), \text{vec}(\mathbf{X}_{t-1}^{(2)} \mathbf{X}_{t-1}^{(2)\text{T}})) = 4\text{Var}(\text{vec}V^\nu) + 2\text{Var}(\text{vec}V^\omega) + \text{Var}(\text{vec}S_{t-1}^\nu) \tag{61}$$

$$\text{Cov}(\text{vec}(\mathbf{X}_t^{(1)} \mathbf{X}_t^{(1)\text{T}}), \text{vec}(\mathbf{X}_{t-s}^{(2)} \mathbf{X}_{t-s}^{(2)\text{T}})) = 4\text{Var}(\text{vec}V^\nu) + 2\text{Var}(\text{vec}V^\omega) \quad \forall t, \forall s \geq 2 \tag{62}$$

These results are obtained by focussing on a general element of a matrix on the left hand side, and then substituting into the left hand sides the definitions (19) and (20), expanding the covariances, substituting representations (12) and (13), and then simplifying the result using known orthogonalities to deduce the general element of the matrices on the right hand side. However, there are several hundred terms in some of the expansions and a computer algebra package was used to ensure the accuracy of the results.

References

- Brown, P., N. Le, and J. Zidek (1994). Inference for a covariance matrix. In P. Freeman and A. Smith (Eds.), *Aspects of Uncertainty: A Tribute to D.V. Lindley*, pp. 77–92. Wiley.
- Farrow, M. and M. Goldstein (1993). Bayes linear methods for grouped multivariate repeated measurement studies with application to crossover trials. *Biometrika* 80(1), 39–59.
- Goldstein, M. (1986). Exchangeable belief structures. *J. Amer. Statist. Ass.* 81, 971–976.
- Goldstein, M. (1988). The data trajectory. In J.-M. Bernardo et al. (Eds.), *Bayesian Statistics 3*, pp. 189–209. Oxford University Press.
- Goldstein, M. (1994). Revising exchangeable beliefs: subjectivist foundations for the inductive argument. In P. Freeman and A. Smith (Eds.), *Aspects of Uncertainty: A Tribute to D. V. Lindley*. Wiley.
- Goldstein, M. and D. Wooff (1995). Bayes linear computation: concepts, implementation and programming environment. *Statistics and Computing*, to appear.
- Leonard, T. and J. Hsu (1992). Bayesian inference for a covariance matrix. *Ann. Statist.* 20, 1669–1696.
- Queen, C. (1994). Using the multiregression dynamic-model to forecast brand sales in a competitive product market. *Statistician* 43:1, 87–98.
- Queen, C., J. Smith, and D. James (1994). Bayesian forecasts in markets with overlapping structures. *International journal of forecasting* 10, 209–233.
- West, M. and P. Harrison (1989). *Bayesian forecasting and dynamic models*. New York: Springer.
- Wilkinson, D. and M. Goldstein (1995). Bayes linear adjustment for variance matrices. In J.-M. Bernardo et al. (Eds.), *Bayesian Statistics 5*, to appear.
- Wooff, D. (1992). [B/D] works. In J.-M. Bernardo et al. (Eds.), *Bayesian Statistics 4*, pp. 851–859. Oxford University Press.